

FIG. 1

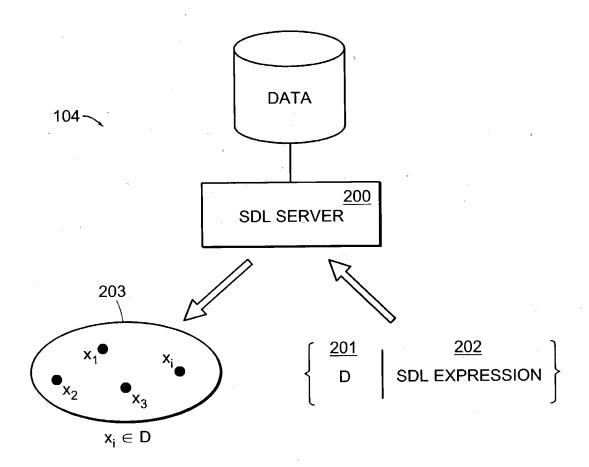


FIG. 2

3/35

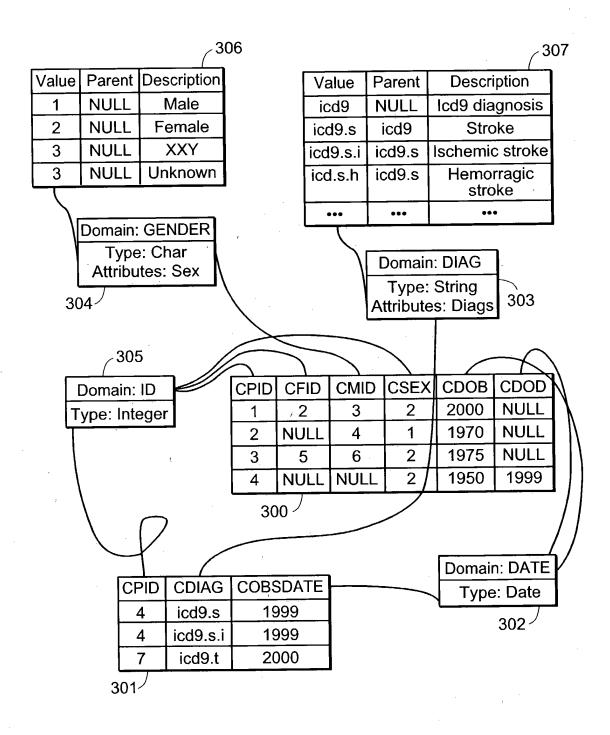


FIG. 3

400 Domain SQLtype Hierarchy Min Max 303-DIAG String **TDiags** NULL **NULL** 304-GENDER Char TSex NULL NULL 302 NULL DATE 1900 2100 Date 305-ID NULL 0 NULL Integer

<sub>/</sub>401

<sub>/</sub>403

Dimension	Domain
PID	ID
FID	ID
MID	ID
SEX	GENDER
DOB	DATE
DOD	DATE
OBSDATE	DATE
DIAG	DIAG

Dimension	Relation	Column_name	Multiplicity
PID	relation1	CPID	Unique
PID	relation2	CPID	Multiple
FID .	relation1	CFID	Multiple
MID	relation1	CMID	Multiple
SEX	relation1	CSEX	Multiple
DOB	relation1	CDOB	Multiple
DOD	relation1	CDOD	Multiple
OBSDATE	relation2	COBSDATE	Multiple
DIAG	relation2	CDIAG	Multiple

402

Relation	SQLrelation	Inclusion_criteria
relation1	TableA	{PID, FID, MID, SEX, DOB, DOD}
relation2	TableB	Sex

5/35

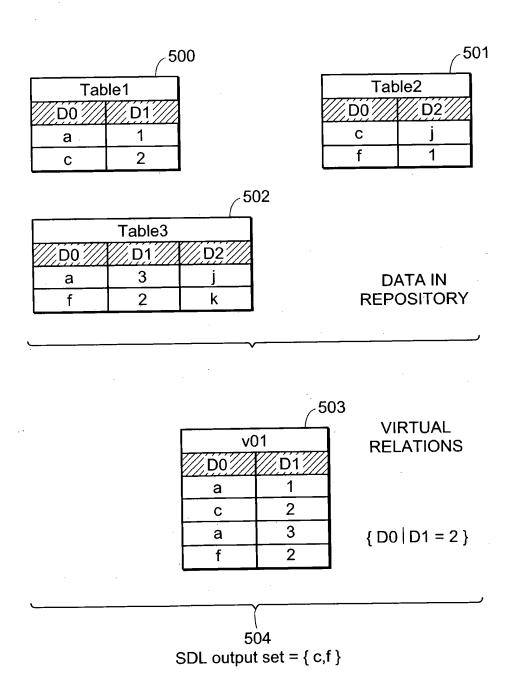


FIG. 5

6/35

**€**600

PID	DIAG	HOSPID	DOCID	DATE
1	icd9.stroke	100	Α	1966
2	icd9.x	200	В	2000

601

HOSPID	NAME	TYPE	REGION	ZIP
100	MGH	TH	Boston	02222
200	BWH	TH	Brookline	02115

602

PID	DATE	DIAG	HOSP.	HOSP. TYPE	HOSP. REGION	HOSP. ZIP
1	1966	icd9.stroke	MGH	TH	Boston	02222
2	2000	icd9.x	BWH	TH	Brookline	02115

603

Relation	SQL relation	Inclusion_criteria
relation1	diags	{PID,DIAG,HOSPID,DOCID}
relation2	hospitals	{HOSPID,NAME,TYPE,REGION, ZIP}
relation3	vdiags	{PID,DIAG,HOSPID,DOCID}, {HOSP.NAME,HOSP.TYPE, HOSP.REGION,HOSP.ZIP}

604

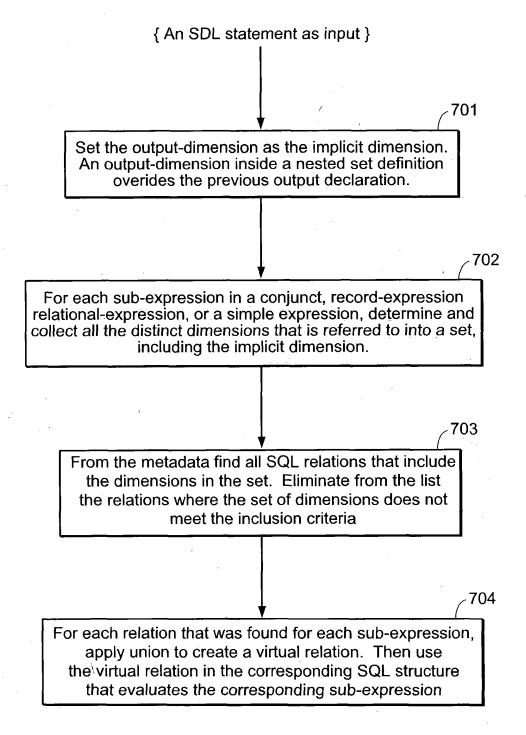
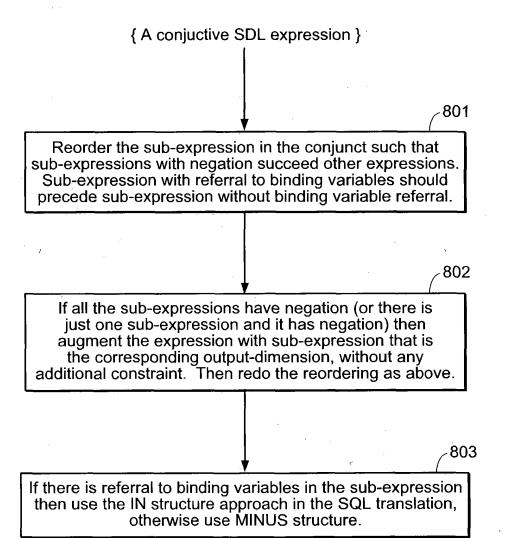


FIG. 7



15-ر

A disjunctive SDL expression of the form (expr\_A) OR (expr\_B)

900

Translate to SQL using the union approach, e.g.:
SELECT DISTINCT output-dim
FROM (SDL2SQL(expr\_A) UNION SDL2SQL(expr\_B)

A conjunctive SDL expression of the form (expr\_A) AND (expr\_B)

901

Translate to SQL using the join approach, e.g.:
SELECT DISTINCT output-dim FROM
(SDL2SQL(expr\_A) vA, SDL2SQL(expr\_B) vB
WHERE vA.output-dim = vB.output-dim

FIG. 9

{ A general SDL expression } Apply OR-distribution continuously, including on record-operators and nested sets, until the expression is in CNF. Brackets determine evaluation order Apply OR-merge on conjuncts that have the same set of dimensions, binding variable referrals and negation structure Reorder terms such that negations succeed terms without negation. Apply negation rewrite procedures. (FIG. 8) Declare conjuncts with referral to binding variables that are not defined within the same conjunct as false Translate each conjunct separately to SQL and apply the union approach to combine disjunctive parts.

FIG. 10

13~

{ An SDL conjunct }

Determine the virtual relations that are required for all the terms in the conjunct

Instead of joining the virtual relations, apply the distribution law on the SQL-relations (e.g. tables) and translate the join of virtual relations into multiple joins (conjuncts) of distinct SQL relations.

Continue with the translation of each join, now using SQL relations instead of the virtual relations, and combine the joins with union.

FIG. 11

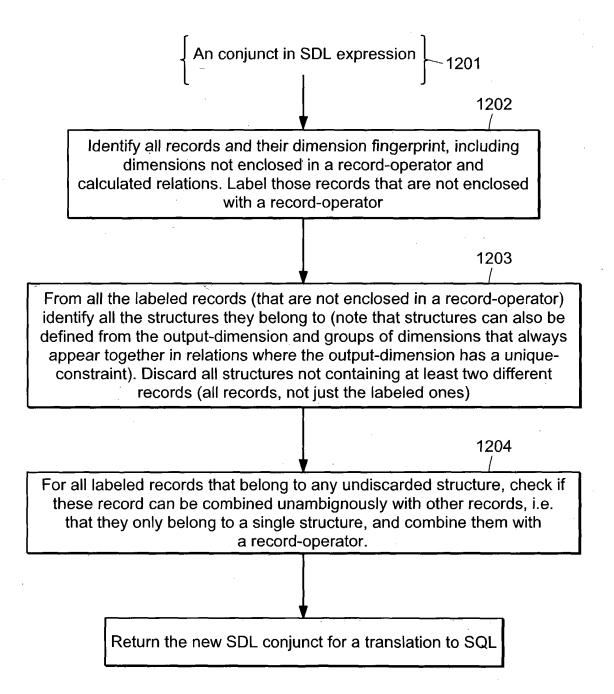


FIG. 12

An SDL statement with inlined SSDL as an input

Scan the statement for set definitions {select...}, SQLV(...) value function or virtual relations [select...] containing SSDL expressions.

1301

For each {select...}, SQLV(...) or [select...] expressions: Call SSDL parser (Figure 14) to parse and generate valid SQL and pass it a common dimension and column alias generator. This may recursively call the SDL parser. Also, collect all common-table expression generated by the SSDL parser.

For the SQL that results from {select...,include it in the combined SQL statement as the SQL code resulting from nested SDL or code within the SDL SIZE function. For the virtual relation that results from [select...] generate (collect) a common-table expression and create an equi-join with the columns representing the output-dimension in the corresponding conjunct.

Return all common-table expressions collected above to be used in a global WITH clause.

Generate SQL for the remainder of the SDL query (as in Figure 10 or Y if it contains SSDL) and return it.

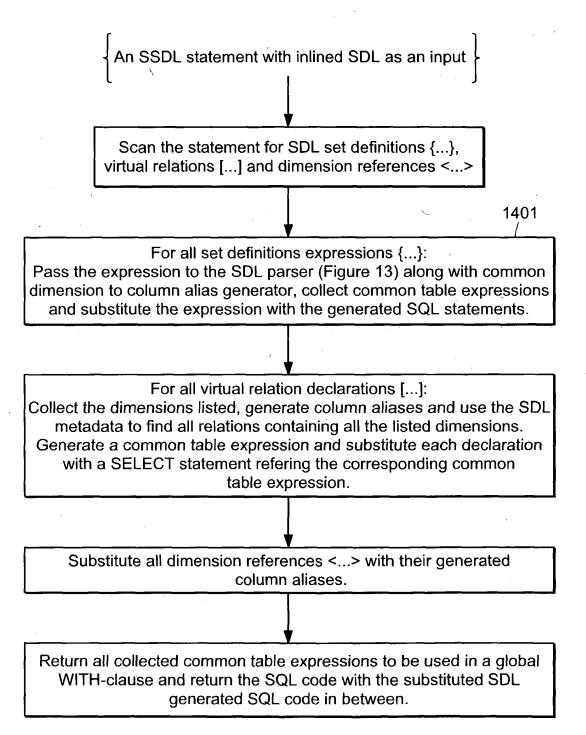
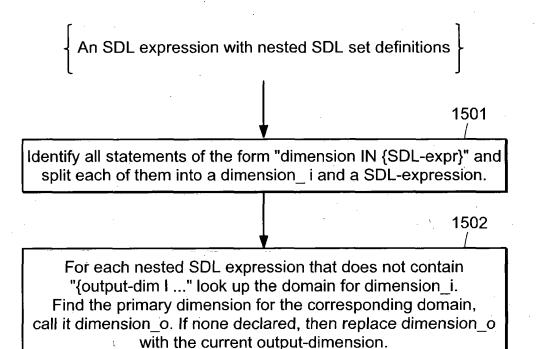


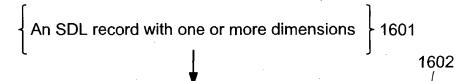
FIG. 14



1503

For each nested SDL expression that did not contain an output-dimension substitute { dimension\_o I SDL-expr } (preferrably with prefix notation) and return the new SDL expression for further SQL translation.

FIG. 15



If any dimension is not recognized as a regular dimension in the metadata, see if it can be a virtual dimension by matching any regular dimension name from left to the corresponding dimension, taking into account prefix from a current output-dimension. Based on the primary dimension of the domain of the dimension for which the match was found, continue this process and collect the dimensions found in each step until the trailing part is recognized as a regular dimension. If the trailing part is not a valid dimension, return an unknown status. Repeat the process for all the dimensions in the record.

1603

For each of the virtual dimensions in the record check to see if the dimensions that were collected are the same (i.e. the same join path) and that the dimensions found from the trailing parts belong to a record. If not, return a flag indicating that the dimensions don't belong to the same record.

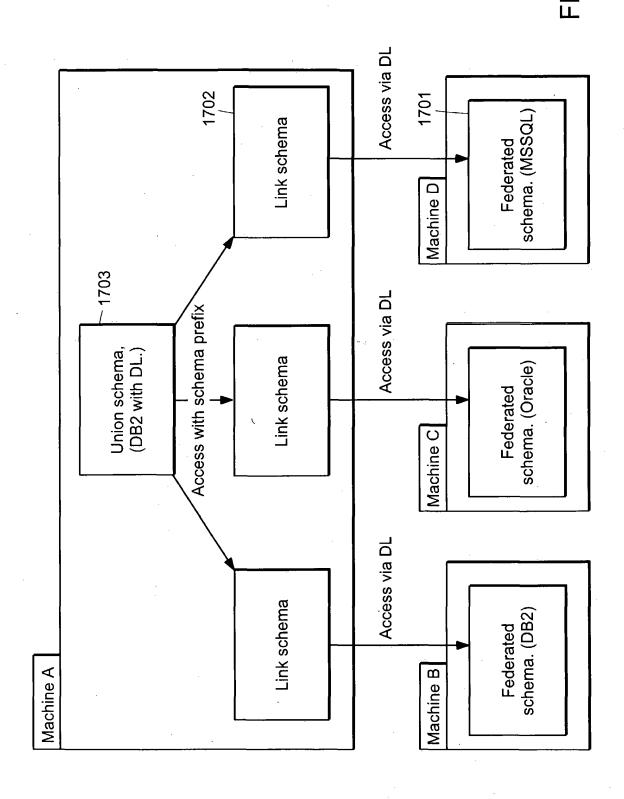
1604

Based on the collected dimensions (the join path) form a virtual relation by a join based on a virtual relation with the output-dimension and the first collected dimension and a virtual relation with the primary dimension indicated by the first collected dimension and the second collected dimension etc., until all virtual relations in the join path have been joined together with an equi-join on the previous collected dimension and the corresponding primary dimension in the next relation. The final virtual relation used in the join is determined by all the dimensions found from the trailing parts in the virtual relations and the primary dimension defined by the prior collected dimension.

1605

Accept the dimensions as virtual dimensions and return a structure to build a common table expression that defines the virtual relations as described in previous step.

17/35



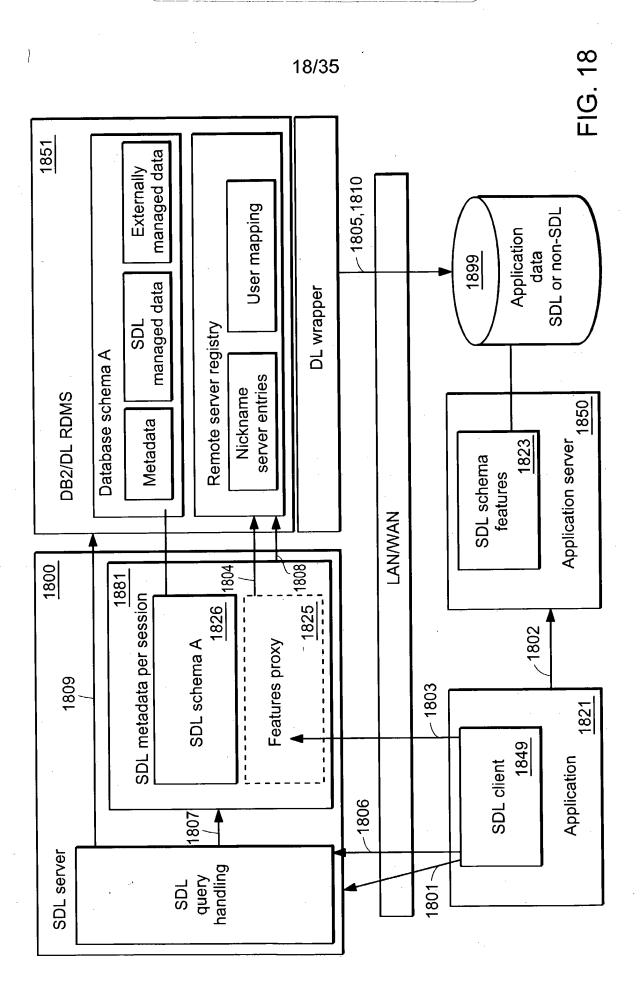


	FIG. 19A	FIG. 19B	FIG. 19		19/35	-1905		FIG. 19A
	SDL Editor[External  Venn  Cutput Spec  Dialog Builder  Resources   [미동를 G한미 대		1903		1904	Intersect	Name         Expression         Date         Count         Dimensions           Query 0         ALT_IDS SYMBOL         Wed Feb 06 17:00:20 GMT 2002         14620         GID           Query 1         ALT_IDS SYMBOL and NOT LOC.CHROM/Wed Feb 06 17:01:00 GMT 2002         47         GID           Query 3         ALT_IDS SYMBOL and LOC.CHROM/         Wed Feb 06 17:02:01 GMT 2002         14573         GID           Query 3         ALT_IDS SYMBOL and LOC.CHROM/         Wed Feb 06 17:02:01 GMT 2002         14573         GID	
图 Query Tool	♀ □ Projects ♀ □ Gene Mining ♀ □ Dimensions	O Mernative gene IDs O Accession O Gene Symbol O Gene Symbol	88	0000	© ☐ Protein Structure and function © ☐ Enzyme reactions © ☐ GeneOntology  1902 ○ ☐ GeneOntology  ○ ☐ Original go id © ☐ Interaction © ☐ Protein motifs © ☐ Sequence Similarity © ☐ Species Homology © ☐ Transcript related infomation	○ ☐ Resources	Related Dimensions  © Gene ID  © Original go id  Umport identifier	

@Query Tool	
o ☐ Projects	SDL Editor   External   Venn   Output Spec   Dialog Builder   Resources
9 ☐ Alternative gene IDs	PF.GO.TERM = : 3673.8150.16265.8219
Accession  Accession	1908
Gene Symbol Alias	
⊙ ☐ Expression	
O Gene ID	
0	
] {	
	-
O-[] Mutations and Polymorphisms	
	Go Terms = : cell death
G-[ Enzyme reactions	
O GeneOntology	
1907 © Diological_process	
φ	
biological_process unkown	
©- (∪) cell communication ©- (∩) cell crowth and/or maintepance	
60	
⊙ ⇔ cell death	
© Chiesus death © C) di Type: Value	Intersect Query
⊕ CD o Name: 3673.8150.16265.8219	
© O pl Description: cell death	Expression Date Count
	ALT_IDS SYMBOL   Wed Feb 06 17:00:20 GMT 2002   14620
⊙ ○ cellular component	ALT IDS SYMBOL and NOT LOC.CHROM wed Feb 06 17:01:00 GMT 2002
adianilas inalas 🗥 🕜	Wed rep to 17:01:27 GMT 2002 01:20 Wed rep to 17:02:01 GMT 2002 14573
Related Dimensions	ALL TIPS STRIBOLE AND LOCK-CHINOLW WED FED ON 17:02:01 GWT 2002 14:513
☐ Gene (D	
Original go id	

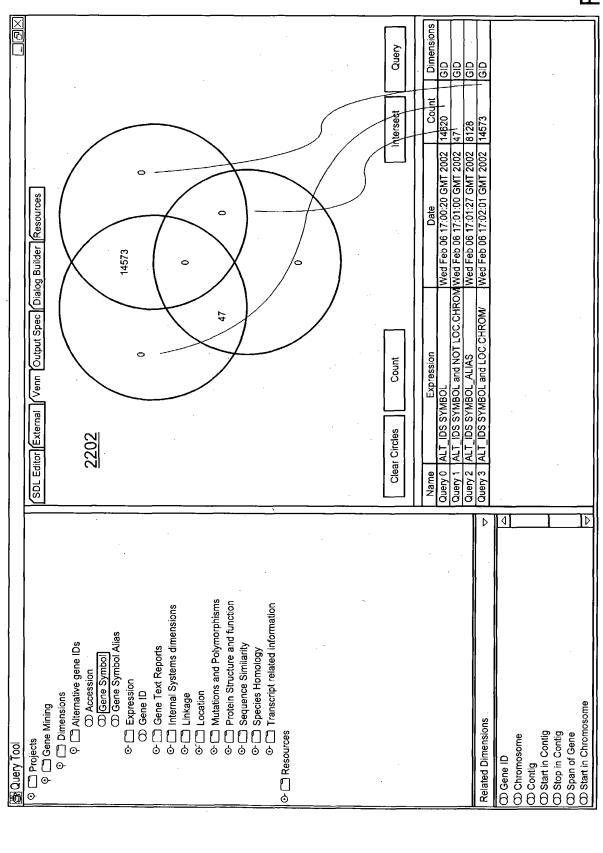
图Query Tool		
⊋ □ Projects	SDL Editor   External   Venn   Output Spec   Dialog Builder   Resources	FIG 20A
	다 있 음 등 대 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로 로	
	{ALT_IDS.SYMBOL and LOC.CHROM }	
2000 O Accession	2002	FIG. 20B
∴ Gene Symbol ○ Gene Symbol Alias	1001	
© □ Expression Company Compan		FIG 20
8 🛛		
○ □ Internal Systems dimensions         ○ □ Internal Systems dimensions		
		1
© ☐ Protein Structure and function	{ Gene Symbol and Chromosome }	<i></i>
][		
©- Transcript related information		
→ ☐ Resources		21 <i>i</i>
		/35
		5
	Intersect	
	Ouev 0 At 1 DS SYMBOL Wed Feb 06 17:00:20 GMT 2002 14620 GID	
	ALT IDS SYMBOL and NOT LOC CHROM Wed Feb 06 17:01:00 GMT 2002   47	
	IDS SYMBOL_ALIAS Wed Feb 06 17:01:27 GMT 2002   8128	
Related Dimensions	ALI IDS STMBOL and LOC.CHROM/	
D Gene ID		
Continuosome 2001		
Start in Contia		
D Stop in Contig		
D Span of Gene		· · · · · · · · · · · · · · · · · · ·
D Start in Chromosome   □		FIG 20A

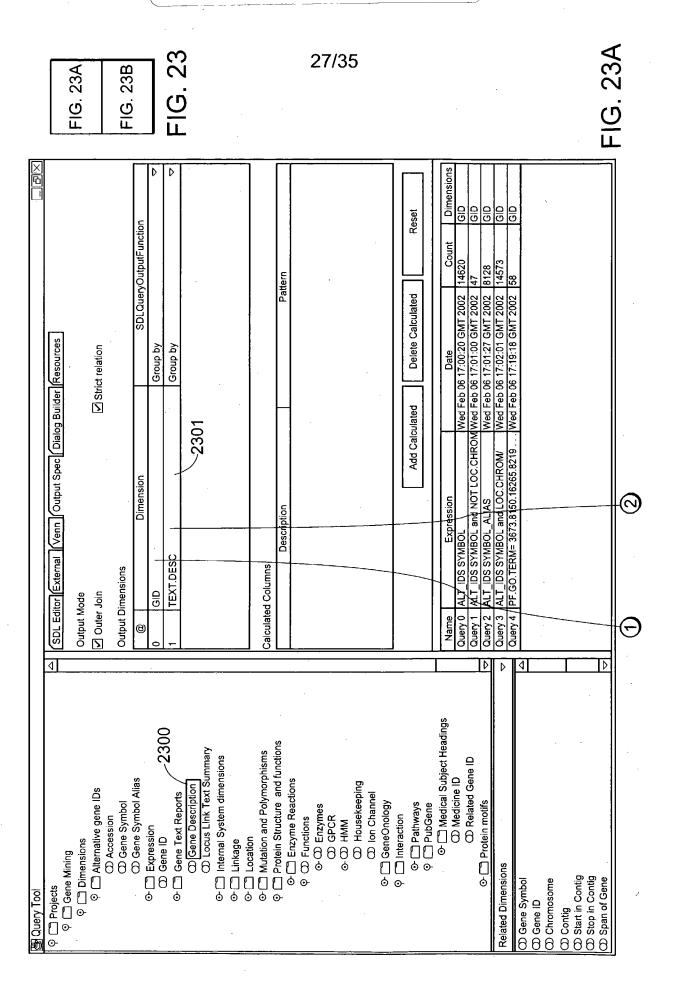
♀ 🗀 Projects	SDL Editor External   Venn   Output Spec   Dialog Builder   Resources
♀ ☐ Gene Mining	
O C Alternative gene IDs	{ ALT_IDS.SYMBOL and LOC.CHROM }
(1) Accession	
Gene Symbol Alias	
⊕ ☐ Expression	
General Sustant Sustan	
] [	
][	
10	
][	
<u>ן</u>	Gene Symbol and Chromosome
O Species Homology	
][	
G Resources	
,	
	Ļ
	Intersect   Query
	Expression Date Count
	ALT   IDS SYMBOL
-	Query 1 ALI IDS SYMBOL and NOT LOC.CHROMWED FED UP 17:01:00 GMT 2002   47   GID
Related Dimensions	ALT_IDS SYMBOL and LOC.CHROM/ Wed Feb 06 17:02:01 GMT 2002 14573
○ Gene ID	
Chromosome	
○ Contig	
C Start in Contig	
☼ Stop in Contig	
Span of Gene	
☼ Start in Chromosome	,

Carery 1001		
	(SDI Editor (External Tylenn Minter Space Minider Resources	
© Cene Mining	The state of the s	FIG 21A
		: : : : :
♀ ☐ Alternative gene IDs	PF.GO.TERM - :3673.8150.16265.8219 and PF.INTER.PUB.PGID	
		2,0
Gene Symbol     Gene	eswold	FIG. 21B
	<u>a</u>	
	т.	
O Gene ID		- LC 071
○- ☐ Gene Text Reports	7.100	
	₩	
⊙ ☐ Linkage		
O 🗀 Location		
O- ☐ Mutations and Polymorphisms		-
→ ☐ Protein Structure and function	Go Terme = 1.01 death and Belated Gane ID	
⊕	כל לפונים כיינים משמים וכיימים מחום ביינים כיינים כ	
◆ ☐ Functions		
⊕ ☐ GeneOntology		
O Interaction		
⊕ ☐ Pathways		23.
O D PubGene		/3
		5
(C) Medicine ID		
CD Related Gene ID		
⊕ ☐ Protein motifs		
⊕ ☐ Sequence Similarity		
⊕ Species Homology	Intersect   Query	
⊙ ☐ Transcript related information		
⊕ ☐ Resources	r	
	ALT IDS SYMBOL	
	ALT IDS SYMBOL and NOT LOC CHROM Wed Feb 06 17:01:00 GMT 2002 47	
	ALT IDS SYMBOL ALIAS   Wed Feb 06 17:01:27 GMT 2002   81:28	
Polated Dimensions	ALT IDS SYMBOL and LOC CHROM/ Wed Feb 06 17:02:01 GMT 2002 14573	
Verated Diffielisions		
Gene ID		
○ Medicine ID		
CD Mesh String		
Mesh Terms     Mesh Terms		
□ Import identifier     □ Import identifier		
		; ; ;
		710 DIA

圖 Query Tool	
Q ☐ Projects	(SDI Editor/External   Venn   Output Special Ballider   Resources
© 🗀 Alternative gene IDs	PF.GO.TERM = :3673,8150.16265,8219 and PF.INTER.PUB.PGID IN { }
(U) Accession	Browse Dimension
○ Gene Symbol Alias	ALT IDS.ACCESSION - Accession
⊕ ☐ Expression	POLY_MUT.MRK.ACCESSION - Accession
<del>0</del> 1	PE:PROT.BEST_PROT_ID - Best Match Protein ID
Õ (	LOC.CHROM - Chromsome
○ ☐ Protein Structure and function	Go Terms = :cell death and Related Gene ID IN {}
⊙- ☐ Pathways	
⊕ ☐ Medical Subject Headings	
Medicine ID	
<ul> <li>✓ Related Gene ID</li> <li>✓ Protein motifs</li> </ul>	
Sequence Similarity	
O Species Homology	Intersect Query
Transcript related information	
○ □ Resources	Name Expression Date Count Dimensions
	ALT_IDS SYMBOL Wed Feb 06 17:00:20 GMT 2002   14620
	ALT IDS SYMBOL and NOT LOC. CHROM Wed Feb 06 17:01:00 GMT 2002   47
	ALT IDS SYMBOL_ALIAS Wed Feb 06 17:01:27 GM   2002   8128
Related Dimensions	Query 3 ALT_IDS SYMBOL and LOC.CHROW Wed Feb 06 17:02:01 GMT 2002   14573   GID
◯ Gene ID	-
Medicine ID	
Mesh Terms	
Umport identifier	

		FIG. 22A			FIG 22B			i	FIG. 2	i - -								2	5/	<b>3</b>	5												•	FIG 22
	SDI Editor (External IVenn Mutaut Spec (Dialog Builder (Respurees)			PF.GO.TERM = :3673.8150.16265.8219 and PF.INTER.PUB.PGIDIN {PF.FUNC = :20.20}				0000	0077					ı	Go Terms = :cell death and Kelated Gene ID IN { Functions = :GFCF }									Intersect		Expression Date Count	ALT IDS SYMBOL Wed Feb 06 17:00:20 GMT 2002 14620	Query 2   ALT IDS SYMBOL ALIAS   Wed Feb 06 17:01:27 GMT 2002   47	ALT IDS SYMBOL and LOC.CHROM/ Wed Feb 06 17:02:01 GMT 2002 14573	2201				
Concert lool	Projects     Projects	© Gene Mining	© 🗂 Dimensions		O Accession	Gene Symbol	_ [		8	O C Gene Text Reports	○ ☐ Internal Systems dimensions	○ □ Linkage	o ☐ Location			○ ○ Functions	⊕	9 GPCR	WWH ⊖ ⊕	○ Housekeeping	O lon Channel		○- ☐ Pathways	© Medical Subject Headings	Cross pricing	Chaire Carle Of Protein adjita	Sequence Similarity	⊙ ☐ Species Homology	Related Dimensions	CD Gene ID CD Medicine ID CD Mesh String	Mesh Terms	O Import identifier		





				. 1			_				_					_								1.	<u> </u>
		2302	DESC	microtubule-associated protein tau  DNA fragmentation factor, 45 kD, albha polypeptide	transforming growth factor, beta 1 APG5 (autophady5, 6, cerevislae)-like	tin ""	Coagulation lactor II (mionibili receptor) (tumor necrosis factor (ligand) superfamily member 8		islet amyloid polypeptide	v-abi Abel Soi i III illi ille leukei III a vilai olicogene i loi loigi. mitogen-activated protein kinase kinase kinase 5	Bruton agammaglobulinemia tyrosine kinase	adenosine A2a receptor	ull E receptor 3 (subtype EF3)	interleukin 2 receptor, alpha	BH3 interacting domain death agonist	PRKC, apoptosis, WT1, regulator	tumor necrosis factor receptor superfamily, member 12 (translocating	xidase	tumor protein p53 (Li-Fraumeni syndrome)	tumor necrosis ractor (ligand) superramily, member o protein tyrosine kinase 2 heta	signal transducer and activator of transcription 1, 91kD	somatostatin receptor 3	insulin-like growth factor 1 receptor	TIA1 cylotoxic granule-associated RNA-binding protein	Abort
(S)	圖 SDL query [58]	(1% 0	ĞIĐ	HS. 101174 Microtubule Hs. 105658 DNA fragme				6	Hs. 142255		76	Hs. 1613				Hs.176090 PRKC, apop				Hs.2007 He 20313 Diotein tyres		2		Hs.239489 TIA1 cylotox	58 rows (Total: 750 msec)

		FIG. 24A		FIG. 24B			FIG 24	-					29/35				( [	FIG. 24A
× 60 1	Venn Output Spec Dialog Builder Resources	· ·	Strict relation—2403		SDLQueryOutputFunction	Group by   □	Group by	Count distinct	2400			Pattern		Add Calculated Delete Calculated Reset	Date   Count   Dimensions   Count   Dimensions   Count   Dimensions   Count   Dimensions   Count   Dimensions   Count   Dimensions   Count   Count   Dimensions   Count   Count   Dimensions   Count   Count   Dimensions   Count   Dimensions	Wed Feb 06 17:02:01 GMT 2002 145/73	2007 1911 10 00 141 2007 100 00 100 100 100 100 100 100 100 1	
-	SDL Editor External Venn Output Sp	Output Mode	☑ Outer Join2402	Output Dimensions	@ Dimension	O GID	1 TEXT.DESC	2 LINK.DIS			Calculated Columns	Description			Name Expression Query 0 ALT IDS SYMBOL Query 1 ALT IDS SYMBOL and NOT LOC  Ouery 2 ALT IDS SYMBOL and NOT LOC		_	
Cuery Tool	© Projects	y ☐ Cerie willing		O Gene Symbol	○ Gene Symbol Alias	G Expression	O Gene Text Reports		<ul> <li>○ Locus Link Text Summary</li> <li>○ □ Internal System dimensions</li> </ul>	♥ Lunkage ⊙-ဤ Disease	Lod Score	2	<ul> <li>○ Location</li> <li>○ Mutations and Polymorphism</li> <li>○ Location Structure and function</li> <li>○ Locations</li> <li>○ Descriptions</li> <li>○ Descriptions</li> <li>○ OFCR</li> </ul>	© C) HMM  C) Housekeeping  C) Ion Channel	][]ዕଡ	Related Dimensions	© Gene ID © Lod Score © Pheotype © Import identifier	

(GID   %1)		
GID	TEXT DE8C	COUNT(DISTINCT LINK DIS
Hs.101174	microtubule-associated protein tau 9	0
Hs.105658	DNA fragmentation factor, 45 kD, alpha polyp 0	2404
Hs.1103	ctor, beta 1	2401
HS.11171 Hs.13400	APG5 (autophagy5, 6, cerevislae)-like 0	
Hs. 128087	actor II (thrombin receptor	
Hs.1313	tumor necrosis factor (ligand) superfamily m 8	
Hs.133089		
Hs.142255	islet amyold polypeptide 6	
Hs.146355	v-abi Aberson murine leukemia viral oncogen 8	
Hs.151988		
Hs.159494	Bruton agammaglobulinemiatyrosine kinase 0	
Hs.1613	adenosine A2a receptor 6	
Hs.170917	prostaglandin E receptor 3 (subtype EP3) 10	
Hs.1722	interleukin 1, alpha	
Hs.1724		
Hs.172894	BH3 interacting domain death agonist	5
Hs.176090		
Hs.177486	amyloid beta (A4) precursor protein (protease 6	
Hs.180338	tumor necrosis factor receptor superfamily, m 9	
Hs.1817	myeloperoxidase 8	
Hs.1846	tumor protein p53 (Li-Fraumeni syndrome) 8	
Hs.2007	tumor necrosis factor (ligand) superfamily, m 9	
Hs.20313	protein tyrosine kinase 2 beta 6	
Hs.21486	signal transducer and activator of transcriptio 8	
Hs.225995	somatostatin receptor 3	
Hs.239176	insulin-like growth factor 1 receptor 4	
Hs.239489	RNA-bindin	
58 rows (Total: 4426 msec)		Abort

FIG. 25B

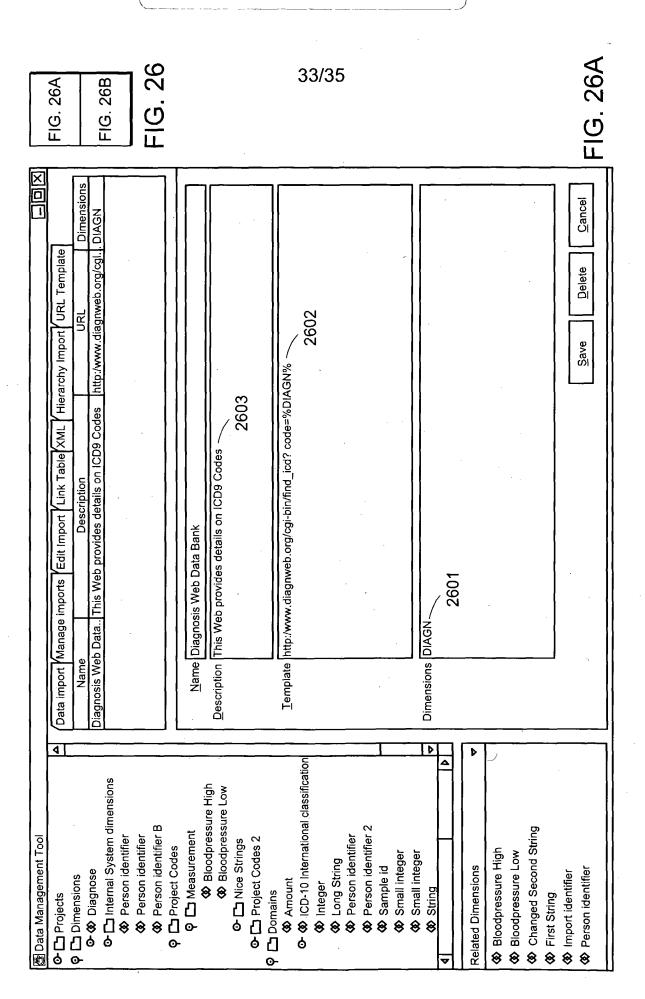
25A

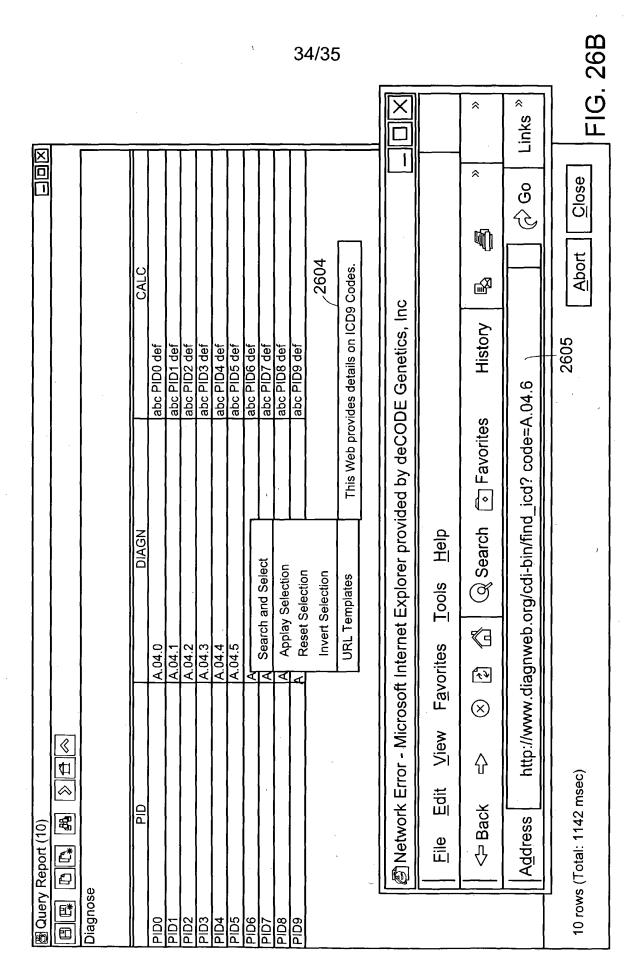
<u>된</u>

FIG. 25

Dimensions 응 Wed Feb 06 17:01:27 GMT 2002 Wed Feb 06 17:00:20 GMT 2002 AND 🗢 IDS SYMBOL and NOT LOC.CHROM Wed Feb 06 17:01:00 GMT 2003 Wed Feb 06 17:02:01 GMT 2003 Output Spec Dialog Builder Resources HTLod.Score Disease D Lodscore calculated with Allegro LINK.LODS Disease association LINK.DIS IDS SYMBOL and LOC.CHROM Test dialog Expression External Venn 2502 G SDL Editor Query 1 Query 2 Query 0 Query 3 M O Medical Subject Headings 2500 ○ ☐ Mutations and Polymorphisms
 ○ ☐ Protein Structure and function Gene Description
 Locus Link Text Summary © 🗀 Internal Systems dimensions 🍳 🗀 Linkage ○ Accession○ Gene Symbol○ Gene Symbol Alias © ☐ Gene Mining ♀ ☐ Dimensions ♀ ☐ Alternative gene IDs O- ☐ Enzyme Reactions ○ ○ HMM○ ○ Housekeeping ○ Gene ID○ □ Gene Text Reports O- ☐ GeneOntology
Q ☐ Interaction
O- ☐ Pathways
Q ☐ PubGene O Ion Channel ∴ Lod Score
 ∴ Phenotype O-☐ Expression ⊙ O Disease O C Location Related Dimensions □ Import identifier Projects MOnery Tool ○ Phenotype O Gene ID Disease

@ Query Tool	
♀ 🗀 Projects	SDI Editor Festernal   Venn   Outhout Spec   Dialog Builder   Resources
♀ ☐ Gene Mining	
. ♀ ☐ Dimensions	
	Gene Mining
O Accession	2501
O Gene Symbol	t007/
•	
C Expression	
$\cup$ 1	O Amd
© ☐ Gene Text Reports	O Ashma O Ashma
Gene Description	0 COPP) 0 Drozestor
☐ Locus Link Text Summary	U Corporation
ш	O Hypertansion
□ Linkage	[D
O Disease     O Disea	
O Lod Score	
① Phenotype	(例 Lest dialog
⊕ ☐ Location	
⊕ ☐ Mutations and Polymorphisms	Test dialog
O- Protein Structure and function	Lodscores calculated with Allegro
O- C Enzyme Reactions	Disease = □ Alzheimers
O (1) Functions	CNA
⊙-O Enzymes	od Smre  = \to  20
O GPCR	
ммн ⊖ ₀	Intersect
○ Housekeeping	] 1
① lon Channel	
○- ☐ GeneOntology	
O-□ Pathways	Expression Date Count
© ☐ PubGene	ALT IDS SYMBOL WELL OF CHICAL WELL FOR 17:00:20 GMT 2002 14620
⊕ C Medical Subject Headings	Query 2 ALT IDS SYMBOL AIMS NOT LOC. CHROWNED FEB 06 17:01:00 GMT 2002 47 GID
Related Dimensions	ALT_IDS SYMBOL and LOC.CHROW Wed Feb 06 17:02:01 GMT 2002 14573
<u>.</u>	7
C Gene ID	126
Disease Shenotine	
© Import identifier	





35/35

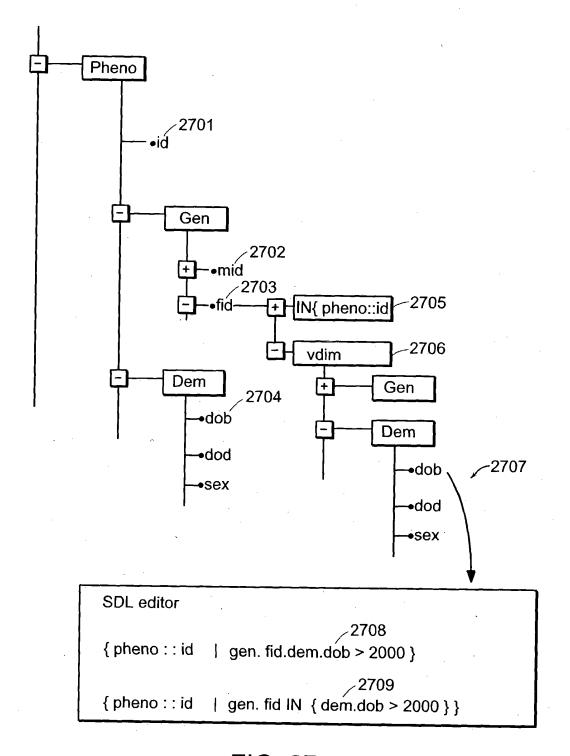


FIG. 27